

Patent Abstract

A relation establishment system for key words in a search engine, which includes a query action database, a browsing action database, and a relation analyzer. The query action database is stored with a first key word, and the associated first searching information, and a second key word, and the associated second searching information. The browsing action database is stored with the first key word and a first browsing information associated with the first searching information, and the second key word and a second browsing information associated with the second searching information. The relation analyzer determines a relation value associated with the first key word and the second key word according to the first searching information, the second searching information, the first browsing information, and the second browsing information.

Liling 12/7/2001

申請日期：	90.12.7	案號：	90130325
類別：	G06F17/30	公 告 本	
(以上各欄由本局填註)			

發明專利說明書

530224

一、 發明名稱 <i>Inventor</i>	中文	搜尋引擎關鍵字之間聯建立系統及方法
	英文	
二、 發明人 <i>Inventor</i>	姓名 (中文)	1. 陳文錡 2. 張履平
	姓名 (英文)	1. Wen-Chih Chen 2. Lu-Ping Chang
	國籍	1. 中華民國 2. 中華民國
	住、居所	1. 台南市金華路二段81巷31-1號 2. 新竹市中山路469巷5號3樓
三、 申請人 <i>Applicant</i>	姓名 (名稱) (中文)	1. 財團法人資訊工業策進會 <i>Information Industry Association</i>
	姓名 (名稱) (英文)	1.
	國籍	1. 中華民國
	住、居所 (事務所)	1. 台北市和平東路二段一〇六號十一樓
	代表人 姓名 (中文)	1. 林逢慶
代表人 姓名 (英文)	1. Feng-Qing Lin (?)	



四、中文發明摘要 (發明之名稱：搜尋引擎關鍵字之關聯建立系統及方法)

一種搜尋引擎之關鍵字關聯建立系統，包括一查詢行為資料庫、一瀏覽行為資料庫及一關聯分析器。查詢行為資料庫中紀錄第一關鍵字及相應之一第一搜尋資訊與一第二關鍵字及相應之一第二搜尋資訊。瀏覽行為資料庫中紀錄第一關鍵字及第一搜尋資訊所相應之一第一瀏覽資訊與第二關鍵字及第二搜尋資訊所相應之一第二瀏覽資訊。關聯分析器則依據第一搜尋資訊、第二搜尋資訊、第一瀏覽資訊、及第二瀏覽資訊，決定第一關鍵字與第二關鍵字所相應之一關聯值。

英文發明摘要 (發明之名稱：)



本案已向

國(地區)申請專利

申請日期

案號

主張優先權

無

有關微生物已寄存於

寄存日期

寄存號碼

無

五、發明說明 (1)

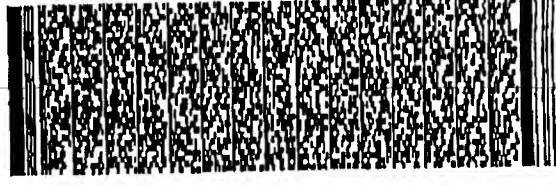
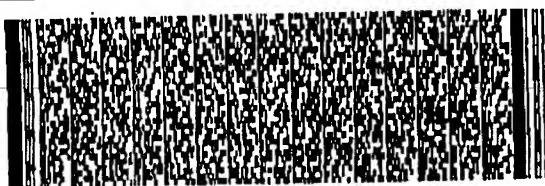
本發明係有關於一種搜尋引擎關鍵字之關聯建立系統及方法，且特別有關於一種可以依據使用者的查詢行為及瀏覽行為來決定關聯詞之關聯建立系統及方法。

在現存的搜尋引擎(Search Engine)中，關聯詞的建立方式主要係依據資訊擷取(Information Retrieval)的方式來進行建立。所謂關聯詞即為在不同關鍵字之間的關聯程度，即當使用者輸入一個關鍵字時，搜尋引擎可以自動查詢或是提供相關的關鍵字給使用者作進一步查詢。然而，若以資訊擷取的方式來自動建立關聯詞，由於每一份搜尋文件並沒有賦予任何實質上意義，而僅透過統計的方法來進行紀錄，因此，往往有正確率(Precision)不高的問題存在。

另一方面，如果以人工的方式建立關聯詞，則必須花費大量的人力與時間成本，而且對於不同領域的關聯詞也不一定能適用，因此，必須重新建立關聯詞而更加浪費人力與時間。此外，由於網際網路世界不斷地有新詞出現，因此，在現存搜尋引擎中，也必須不斷地將新詞建立關聯，以維持整體搜尋引擎之關聯正確率。

有鑑於此，本發明之主要目的為提供一種可以依據使用者的查詢行為及瀏覽行為來決定關聯詞之關聯建立系統及方法。

為了達成本發明之上述目的，可藉由本發明所提供之搜尋引擎之關鍵字關聯建立系統來達成。依據本發明實施例，搜尋引擎之關鍵字關聯建立系統包括一查詢行為資料



五、發明說明 (2)

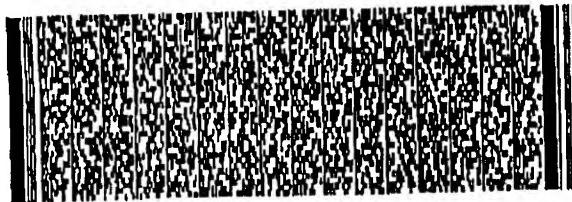
庫、一瀏覽行為資料庫及一關聯分析器。

查詢行為資料庫中紀錄一第一關鍵字及相應之一第一搜尋資訊與一第二關鍵字及相應之一第二搜尋資訊。瀏覽行為資料庫中紀錄第一關鍵字及第一搜尋資訊所相應之一第一瀏覽資訊與第二關鍵字及第二搜尋資訊所相應之一第二瀏覽資訊。關聯分析器則依據第一搜尋資訊、第二搜尋資訊、第一瀏覽資訊、及第三瀏覽資訊，決定第一關鍵字與第二關鍵字所相應之一關聯值。

此外，本發明亦提出一種搜尋引擎之關鍵字關聯建立方法。首先，紀錄一第一關鍵字及相應之一第一搜尋資訊與一第二關鍵字及相應之一第二搜尋資訊。然後，紀錄第一關鍵字及第一搜尋資訊所相應之一第一瀏覽資訊與第二關鍵字及第二搜尋資訊所相應之一第二瀏覽資訊。最後，依據第一搜尋資訊、第二搜尋資訊、第一瀏覽資訊、及第二瀏覽資訊，決定第一關鍵字與第二關鍵字所相應之一關聯值。

依據本發明之一型態，第一搜尋資訊與第二搜尋資訊分別包括一第一文件集合與一第二文件集合，且關聯值包含一搜尋關聯值。其中，搜尋關聯值係依據第一文件集合與第二文件集合中出現之相同文件的數目來決定。

依據本發明之另一型態，關聯值更可以包含一瀏覽關聯值，且第一瀏覽資訊與第二瀏覽資訊，分別包括相應第一文件集合與第二文件集合之一第一瀏覽時間集合與一二瀏覽時間集合。其中，瀏覽關聯值係依據第一文件集合



五、發明說明 (3)

與第二文件集合中每一相同文件之重要性來決定。而每一相同文件之重要性係依據每一相同文件所分別相應第一瀏覽時間集合與第二瀏覽時間集合中之一第一瀏覽時間與一第二瀏覽時間來決定。

圖式簡單說明

為使本發明之上述目的、特徵和優點能更明顯易懂，下文特舉實施例，並配合所附圖示，詳細說明如下：

第1圖為一示意圖係顯示依據本發明實施例之搜尋引擎之關鍵字關聯建立系統之系統架構。

第2圖為一流程圖係顯示依據本發明實施例之搜尋引擎之關鍵字關聯建立方法之操作流程。

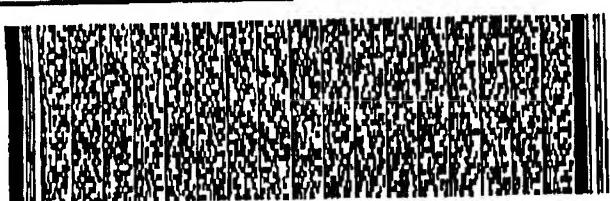
符號說明

- 10~搜尋引擎；
- 11~查詢行為資料庫；
- 12~瀏覽行為資料庫；
- 13~關聯分析器；
- 14~關聯詞資料庫；
- S20、…、S28~操作步驟。

實施例

接下來，第1圖顯示依據本發明實施例之搜尋引擎之關鍵字關聯建立系統之系統架構，參考第1圖，本發明實施例之系統架構將說明於下。

依據本發明實施例之搜尋引擎之關鍵字關聯建立系統包括一搜尋引擎10、查詢行為資料庫11、瀏覽行為資料庫



五、發明說明 (4)

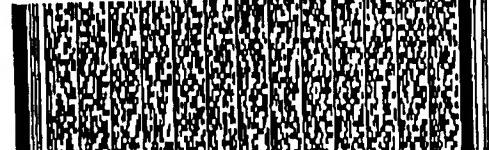
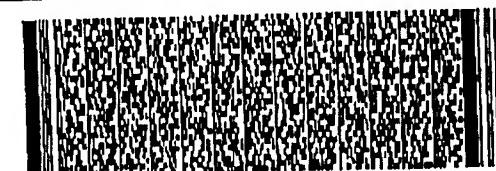
12、關聯分析器13、以及一關聯詞資料庫14。

搜尋引擎10係用以接收使用者輸入之關鍵字，而依照輸入之關鍵字進行查詢。假設使用者輸入一第一關鍵字與一第二關鍵字，則藉由搜尋引擎10的查詢可以分別得到相應之第一搜尋資訊與第二搜尋資訊。其中，第一搜尋資訊與第二搜尋資訊分別包括一第一文件集合與一第二文件集合，換句話說，第一文件集合與第二文件集合分別包括了多篇與第一關鍵字與第二關鍵字相關的文件。

查詢行為資料庫11中則紀錄使用者輸入之第一關鍵字及相應之第一搜尋資訊與第二關鍵字及相應之第二搜尋資訊。瀏覽行為資料庫12中則紀錄第一關鍵字及第一搜尋資訊所相應之第一瀏覽資訊與第二關鍵字及第二搜尋資訊所相應之第二瀏覽資訊。其中，第一瀏覽資訊與第二瀏覽資訊，分別包括相應第一文件集合與第二文件集合之一第一瀏覽時間集合與一第二瀏覽時間集合。第一瀏覽時間集合與第二瀏覽時間集合分別紀錄使用者針對第一文件集合與第二文件集合中不同文件的瀏覽時間。

關聯分析器13則依據第一搜尋資訊、第二搜尋資訊、第一瀏覽資訊、以及第二瀏覽資訊，決定第一關鍵字與第二關鍵字所相應之一關聯值。關聯詞資料庫14則用來紀錄此第一關鍵字、第二關鍵字及相應之關聯值。其中，關聯值包含一搜尋關聯值與一瀏覽關聯值，主要係依據使用者的查詢行為與瀏覽行為來分別決定。

接下來，關聯分析器13如何依據第一搜尋資訊、第二



五、發明說明 (5)

搜尋資訊、第一瀏覽資訊、以及第二瀏覽資訊，來決定第一關鍵字與第二關鍵字所相應之關聯值將進行說明如下。

首先，相應於第一關鍵字與第二關鍵字之關聯值中之搜尋關聯值可以以下列方程式計算：

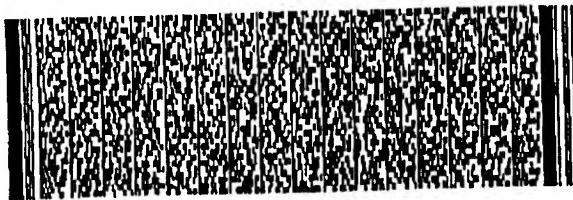
$$\text{搜尋關聯值 } L_2 = \frac{2 \times S}{N_1 + N_2},$$

其中， N_1 表示第一關鍵字所相應之第一文件集合中所包含的文件數目， N_2 表示第二關鍵字所相應之第二文件集合中所包含的文件數目，且 S 表示第一文件集合與第二文件集合中相同文件的數目。因此，搜尋關聯值可以依據第一文件集合與第二文件集合中出現之相同文件的數目來決定。

另一方面，相應於第一關鍵字與第二關鍵字之關聯值中之瀏覽關聯值可以以下列方程式計算：

$$\text{瀏覽關聯值 } L_2 = \frac{\sum_{i=1}^S W_i}{S}, \text{ 且 } W_i = \frac{T_i}{\frac{\sum_{j=1}^S T_j}{P}},$$

其中， W_i 表示 S 篇相同的文件中，文件 i 對於使用者的重要性。 P 表示 S 篇相同的文件中使用者瀏覽的文件數目， T_i 表示使用者瀏覽文件 i 所花費的時間。換言之，所謂文件的重要性係依據此文件相對於其他相同文件中被使用者瀏覽之文件所花費的瀏覽時間而決定。



五、發明說明 (6)

之後，可以利用搜尋關聯值與瀏覽關聯值計算第一關鍵字與第二關鍵字所相應之關聯值：

關聯值_{1,2} = $\alpha \times$ 搜尋關聯值_{1,2} + (1 - α) \times 瀏覽關聯值_{1,2}，其中， $0 \leq \alpha \leq 1$ ，且 α 值可以依據不同的領域或應用而有所調整。

此外，本發明亦可依據不同使用者對於關鍵字的搜尋與文件的瀏覽行為來相互進行關鍵字間的關聯。

假設使用者A與使用者B分別輸入第一關鍵字與第二關鍵字進行查詢。第一關鍵字所相應之第一文件集合中包含了 N_1 篇文件，第二關鍵字所相應之第二文件集合中包含了 N_2 篇文件，且 S 表示第一文件集合與第二文件集合中相同文件的數目。因此，相應於第一關鍵字與第二關鍵字之間關聯值中之搜尋關聯值可以以下列方程式計算：

$$\text{搜尋關聯值}_{1,2} = \frac{2 \times S}{N_1 + N_2},$$

另一方面，相應於第一關鍵字與第二關鍵字之間關聯值中之瀏覽關聯值可以以下列方程式計算：

$$\text{瀏覽關聯值}_{1,2} = \frac{\sum_{k=1}^s \left(\frac{\sum_{i=1}^n W_{i,k}}{n} \right)}{S}, \text{ 且 } W_{i,k} = \frac{T_{i,k}}{\frac{\sum_{j=1}^p T_{j,k}}{\sum_{j=1}^p A_j} P_k}$$



五、發明說明 (7)

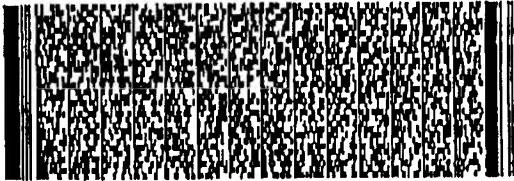
其中， $W_{i,k}$ 表示 S 篇相同的文件中，文件 i 對於使用者 k 的重要性。 P_k 表示 S 篇相同的文件中使用者 k 瀏覽的文件數目， $T_{i,k}$ 表示使用者 k 瀏覽文件 i 所花費的時間， A_i 表示所有使用者瀏覽文件 i 所花費時間的平均。在此情況中，瀏覽行為資料庫 12 中可以依據不同使用者分別紀錄其瀏覽資訊，因此， A_i 可以透過簡單的運算求得。

相同地，可以利用搜尋關聯值與瀏覽關聯值計算第一關鍵字與第二關鍵字所相應之關聯值：

關聯值_{1,2} = $\alpha \times$ 搜尋關聯值_{1,2} + (1 - α) × 瀏覽關聯值_{1,2}，其中， $0 \leq \alpha \leq 1$ ，且 α 值可以依據不同的領域或應用而有所調整。

接下來，第 2 圖顯示依據本發明實施例之搜尋引擎之關鍵字關聯建立方法之操作流程，同時參考第 1 圖與第 2 圖，本發明實施例之操作流程將說明於下。

首先，如步驟 S20，使用者利用搜尋引擎 10 查詢第一關鍵字與第二關鍵字，從而分別得到相應第一關鍵字與第二關鍵字之第一搜尋資訊與第二搜尋資訊。然後，如步驟 S22，將第一關鍵字及相應之第一搜尋資訊與第二關鍵字及相應之第二搜尋資訊紀錄於查詢行為資料庫 11 中。其中，第一搜尋資訊與第二搜尋資訊分別包括一第一文件集合與一第二文件集合，也就是說，第一文件集合與第二文件集合分別包括了多篇與第一關鍵字與第二關鍵字相關的文件。



五、發明說明 (8)

並如步驟S24，將使用者對於第一搜尋資訊之瀏覽行為(第一瀏覽資訊)與第二搜尋資訊之瀏覽行為(第二瀏覽資訊)紀錄於瀏覽行為資料庫12中。其中，第一瀏覽資訊與第二瀏覽資訊，分別包括相應第一文件集合與第二文件集合之一第一瀏覽時間集合與一第二瀏覽時間集合。第一瀏覽時間集合與第二瀏覽時間集合分別紀錄使用者針對第一文件集合與第二文件集合中不同文件的瀏覽時間。

接著，如步驟S26，關聯分析器13依據儲存於查詢行為資料庫11中之第一搜尋資訊及第二搜尋資訊與儲存於瀏覽行為資料庫12中之第一瀏覽資訊及第二瀏覽資訊，決定第一關鍵字與第二關鍵字所相應之關聯值。

相同地，關聯值包含一搜尋關聯值與一瀏覽關聯值。其中，搜尋關聯值係依據第一文件集合與第二文件集合中出現之相同文件的數目來決定。瀏覽關聯值係依據第一文件集合與第二文件集合中每一相同文件之重要性來決定。而每一相同文件之重要性係依據每一相同文件所分別相應第一瀏覽時間集合與第二瀏覽時間集合中之一第一瀏覽時間與一第二瀏覽時間來決定。

最後，如步驟S28，將第一關鍵字、第二關鍵字及相應之關聯值紀錄於關聯詞資料庫14之中。

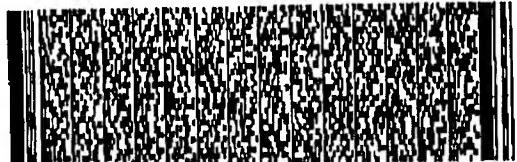
因此，當使用者輸入一關鍵字進行查詢時，搜尋引擎可以依據關聯詞資料庫14中紀錄之相應關鍵字之關聯值來搜尋關聯性較高之關鍵詞來自動查詢，或將關聯性較高之關鍵詞提供給使用者自行查詢。



五、發明說明 (9)

因此，藉由本發明所提出之搜尋引擎之關鍵字關聯建立系統及方法，可以依據使用者或是不同使用者之間的查詢行為及瀏覽行為來決定關聯詞之關聯。

雖然本發明已以較佳實施例揭露如上，然其並非用以限定本發明，任何熟悉此項技藝者，在不脫離本發明之精神和範圍內，當可做些許更動與潤飾，因此本發明之保護範圍當視後附之申請專利範圍所界定者為準。



六、申請專利範圍

1. 一種搜尋引擎之關鍵字關聯建立系統，包括：

一查詢行為資料庫，紀錄一第一關鍵字及相應之一第一搜尋資訊與一第二關鍵字及相應之一第二搜尋資訊；

一瀏覽行為資料庫，紀錄該第一關鍵字及該第一搜尋資訊所相應之一第一瀏覽資訊與該第二關鍵字及該第二搜尋資訊所相應之一第二瀏覽資訊；以及

一關聯分析器，依據該第一搜尋資訊、該第二搜尋資訊、該第一瀏覽資訊、及該第二瀏覽資訊，決定該第一關鍵字與該第二關鍵字所相應之一關聯值。

2. 如申請專利第1項所述之系統，更包括一搜尋引擎，用以查詢該第一關鍵字與該第二關鍵字，從而分別得到相應之該第一搜尋資訊與該第二搜尋資訊。

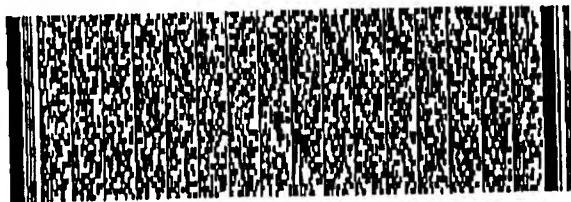
3. 如申請專利第1項所述之系統，更包括一關聯詞資料庫，用以紀錄該第一關鍵字、該第二關鍵字及相應之該關聯值。

4. 如申請專利第1項所述之系統，其中該關聯值包含一搜尋關聯值。

5. 如申請專利第4項所述之系統，其中該第一搜尋資訊與該第二搜尋資訊，分別包括一第一文件集合與一第二文件集合。

6. 如申請專利第5項所述之系統，其中該關聯分析器係依據該第一文件集合與該第二文件集合中出現之相同文件的數目來決定該搜尋關聯值。

7. 如申請專利第6項所述之系統，其中該關聯值更包



六、申請專利範圍

含一瀏覽關聯值。

8. 如申請專利第7項所述之系統，其中該關聯分析器係依據該第一文件集合與該第二文件集合中每一該等相同文件之重要性來決定該瀏覽關聯值。

9. 如申請專利第8項所述之系統，其中該第一瀏覽資訊與該第二瀏覽資訊，分別包括相應該第一文件集合與該第二文件集合之一第一瀏覽時間集合與一第二瀏覽時間集合。

10. 如申請專利第9項所述之系統，其中每一該等相同文件之重要性係依據每一該等相同文件所分別相應該第一瀏覽時間集合與該第二瀏覽時間集合中之一第一瀏覽時間與一第二瀏覽時間來決定。

11. 一種搜尋引擎之關鍵字關聯建立方法，包括下列步驟：

紀錄一第一關鍵字及相應之一第一搜尋資訊與一第二關鍵字及相應之一第二搜尋資訊；

紀錄該第一搜尋資訊所相應之一第一瀏覽資訊與該第二搜尋資訊所相應之一第二瀏覽資訊；以及

依據該第一搜尋資訊、該第二搜尋資訊、該第一瀏覽資訊、及該第二瀏覽資訊，決定該第一關鍵字與該第二關鍵字所相應之一關聯值。

12. 如申請專利第11項所述之方法，更包括查詢該第一關鍵字與該第二關鍵字，從而分別得到相應之該第一搜尋資訊與該第二搜尋資訊。



六、申請專利範圍

13. 如申請專利第11項所述之方法，更包括紀錄該第一關鍵字、該第二關鍵字及相應之該關聯值於一關聯詞資料庫中。

14. 如申請專利第11項所述之方法，其中該關聯值包含一搜尋關聯值。

15. 如申請專利第14項所述之方法，其中該第一搜尋資訊與該第二搜尋資訊，分別包括一第一文件集合與一第二文件集合。

16. 如申請專利第15項所述之方法，其中該搜尋關聯值係依據該第一文件集合與該第二文件集合中出現之相同文件的數目來決定。

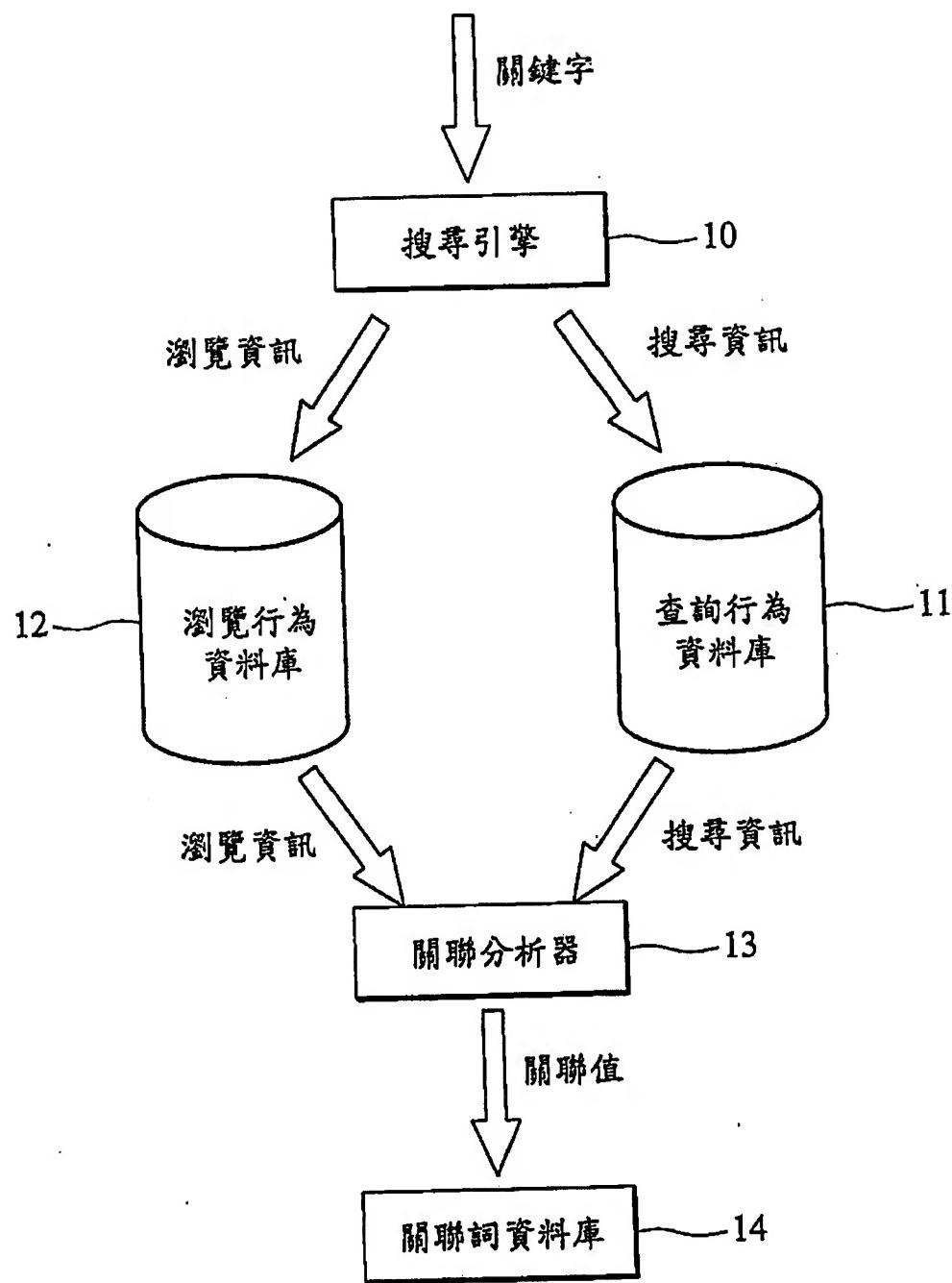
17. 如申請專利第16項所述之方法，其中該關聯值更包含一瀏覽關聯值。

18. 如申請專利第17項所述之方法，其中該瀏覽關聯值係依據該第一文件集合與該第二文件集合中每一該等相同文件之重要性來決定。

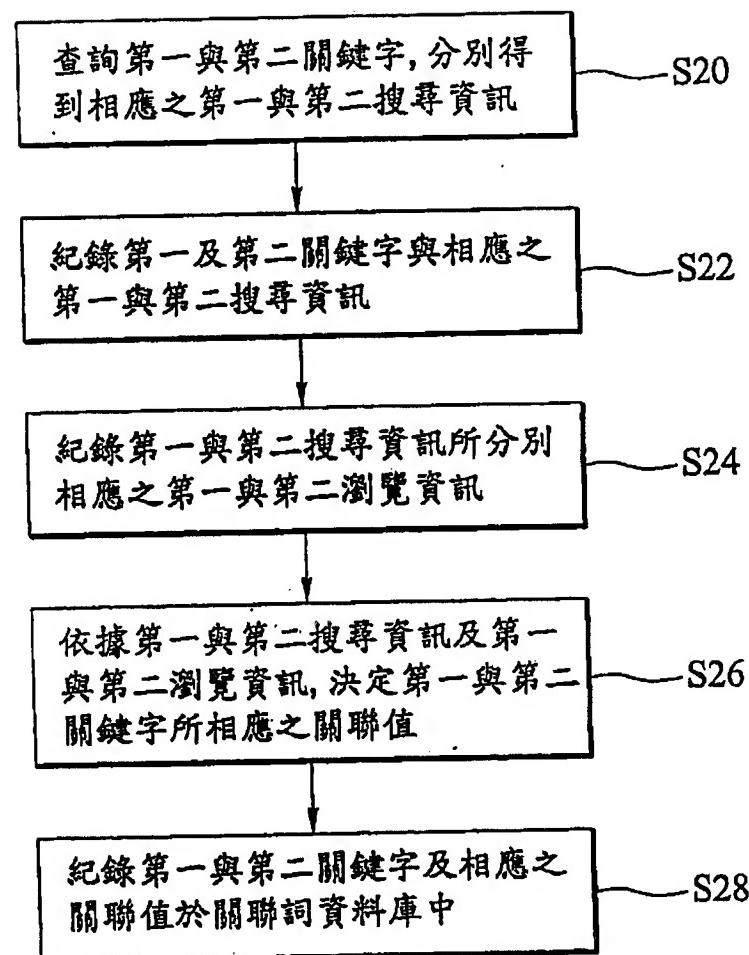
19. 如申請專利第18項所述之方法，其中該第一瀏覽資訊與該第二瀏覽資訊，分別包括相應該第一文件集合與該第二文件集合之一第一瀏覽時間集合與一第二瀏覽時間集合。

20. 如申請專利第19項所述之方法，其中每一該等相同文件之重要性係依據每一該等相同文件所分別相應該第一瀏覽時間集合與該第二瀏覽時間集合中之一第一瀏覽時間與一第二瀏覽時間來決定。





第 1 圖



第 2 圖